AUDITORY NAVIGATION CUES FOR A SMALL 2-D GRID: A CASE STUDY OF THE MEMORY GAME

Jarmo Hiipakka, Gaëtan Lorho, and Jukka Holm

Nokia Research Center, Speech and Audio Systems Laboratory
P.O. Box 407, FIN-00045 Nokia Group, Finland
[jarmo.hiipakka, gaetan.lorho, jukka.a.holm]@nokia.com

ABSTRACT

This paper presents an investigation on auditory navigation in a two-dimensional grid. A sonification of the memory game is proposed and utilized for evaluating user's ability to interact with objects in a small 2-D grid based on sound only. After a short review of the basic game, we start from the idea of replacing each picture under a card by a sound. Subsequently, we present the necessary sonification steps to obtain an eyes-free version of the game. We also propose some slight modifications to the rules of the game to make this audio version more attractive for gaming. This game is used as a case study for auditory user interface testing focusing on auditory navigation in the grid.

1. INTRODUCTION

The memory game or concentration is a well-known board game to develop abilities like concentration, matching and memory skills while playing. The traditional board game, first released in 1966 [1], consisted of 72 picture cards coming in pairs to be laid on a grid. Besides the traditional board game, there are numerous versions of the memory game such as computer, web-based or mobile phone implementations.

In this game, the players try to find identical images randomly scattered in a grid of cards. The user turns two cards at a time to see what's underneath. If the two pictures are identical, the pair is considered solved, and it stays shown. The game ends when all identical pairs are matched.

The memory game was selected as a case study for auditory navigation testing because it provides an easy way to assess a relatively complex user interaction from an intuitive task. Indeed, the game itself is familiar to all subjects, so the participants readily understand the task without long training.

1.1. The audio memory game

The basic idea of the audio memory game is to replace the images in the cards by sounds. The task of the player is then to memorize and match sounds that are either identical or related to each other [2],[3]. Naturally, the matching between images and sounds can also be considered [4].

2. SONIFICATION NEEDS FOR THE GAME

A common feature in the audio memory games presented previously is their utilization of spatial memory for card matching. The grid of cards is presented in the visual domain, and the selection of the cards requires seeing the grid. Therefore, none of these games could be played in a completely eyes-free mode. Our approach is to add audio features to the basic audio memory game until the game can be played in eyes-free situations. First, simple beeps can be added to confirm user action. However, more information can be conveyed to the player with feedback sounds that depend on the cursor position in the grid.

Sonifying the memory game requires the audio representation of a relatively large number of parameters. The game is therefore an interesting application for sonification design and study. The important aspects in developing the sonification for the audio-only memory game are:

- The sonification of the user's position in the grid
- The size of the grid
- The status of the individual cards (found or hidden)
- The number of pairs already found (or remaining)

All this information is readily available to the user, when playing the game with a visual representation of the cards. For an audio-only version, we have to find ways to convey the information, or decide that this information is not necessary.

Notice that the sound design in this case study is aimed at a single player game, in which a card clicked by the user will turn over and a pair will be formed with the next card selected. If the two sounds do not match, then the second card will remain open and the user can search for a new card. This is slightly different from the original idea of the game with two or more players, in which two cards are turned over simultaneously to find pairs.

2.1. Position in the grid

For eyes-free playing of the memory game, knowledge about the current position in the grid of cards is very important. Indeed, the user must be able to find his way back to the sounds he has already heard. For efficiency, the user should be able to easily reproduce the position; so absolute position information is preferred over relative information. In our implementation, the user input is performed via the directional 5-way joystick in the selected test device; therefore, it is evident that perceptual mapping for this 2-D type of grid will be spatial. An ideal sonification should therefore fully support this mapping.

Both mono and stereo output are considered in this study. The mono solution is better suited for reproduction on a loudspeaker or a monophonic headset in a mobile device. However, when headphones are used, stereo reproduction allows stereo panning and 3-D audio to be utilized for spatial presentation of the grid sounds.

For a related problem of presenting the user's current position in a 2-D tabular numeric data display, Ramloll et al. [5] used a solution, in which both horizontal and vertical positions were presented using stereophonic panning. When moving down a column, the first item is presented at extreme left, the last item at extreme right, and the rest in between those. The same scheme is used for horizontal navigation. The justification for this choice was that it is not strictly necessary to present information about the dimension of the grid along which there is no change in position.

2.2. The size of the grid

In the visual domain, the user is constantly aware of the grid size and the relative position in the grid. However, presenting this information with audio is less intuitive. When sonifying the grid, an interaction scheme has to be chosen for notifying the grid boundaries to the user, i.e., deciding on what happens if the user tries to move beyond the first or last row or column of the grid.

One option is to let the position wrap around and just sonify the new position. The user should then be able to know that he just crossed the boundary of the grid. Another option is to prohibit the movement over the boundaries, and provide feedback that indicates that the user cannot go any further.

2.3. Status of the individual cards

As the game advances, the grid will be filled with pairs that have already been matched. As these cards are not part of the problem anymore, a means to distinguish them from the unsolved cards should be included in the sonification. This difference can be implemented as a modification, e.g., filtering or change in timbre, to the normal navigation sounds. The notification could also be an added component in the sounds either for the already matched cards or for the unmatched ones.

2.4. Number of pairs already found

As the game comes near to its end, it would be nice to tell the user he or she is close to the solution. In the visual memory games, this information is often constantly present on the screen in numeric format, and even if this is not the case, it is easy to see the number of pairs not found, as the number of pairs decreases.

3. SONIFICATION DESIGN AND GAME IMPLEMENTATION

A small grid was employed for this experiment with a size of 5 columns and 4 rows. Both a monophonic and a stereophonic version of the sonified memory game were implemented for the test. For the monophonic version, the sonification idea is based on using rhythmic patterns to indicate the row and change in pitch to indicate the column. In the stereophonic version, the sonification is designed using rhythmic patterns and stereophonic panning augmented using 3-D sound techniques for horizontal navigation. In both cases, the status of the card (found/unfound) is notified in the navigation sounds by loudness and timbre alteration.

All the sound samples used for navigation sonification need to be fairly short to allow fast navigation. We used rather clearly pitched, yet percussive synthesized sounds.

3.1. Position sonification for monophonic reproduction

In the sonification for monophonic reproduction, we use the number of sound events to present the current row, i.e., one click or beep for the first row, two for the second, and so on. For the small grid used in the memory game, we found this simple scheme to work very well. For a larger grid, however, it is increasingly difficult to pick out the number of clicks; a different scheme would therefore be needed in such cases.

Our scheme for presenting the current column is based on increasing pitch when the user moves to the left. The problem with the pitch-based cues is that for most people, pitch is relative, and does not provide absolute position information. Therefore, we may expect a difference in navigation accuracy between the absolute representation of the rows and the relative presentation of the columns.

3.2. Position sonification for stereophonic reproduction

3-D sound has been used in different types of auditory displays where spatial sound separation is desired. HRTF-processed sounds are normally heard outside the head when replayed over headphones and can be placed virtually at any position around the listener. However, limitations such as front-back confusions and elevation problems are common with non-individual HRTFs. Spatial sound presentation is well suited for the small grid considered here, which only contains a limited number of sound positions. Displaying sound items along the horizontal plane gives us only one dimension to work with, but it ensures that the spatial sound presentation will work for any user. From previous experiments [6], we know that the users, under certain conditions, can detect five discrete positions from left to right.

Fig. 1 presents three alternatives for spatial sound separation in this restricted case. Standard stereo amplitude panning (left picture) allows only sound lateralization, with a very unpleasant listening experience for signals played at the extreme left or right. With HRTFs (middle picture), the perceptual effect is more natural with a clear externalization at the sides. However, spatial discrimination seems to be less efficient than in the stereo case for mid-positions (i.e. 320º and 40º azimuth), as found in [6]. Therefore, we used a combination of standard amplitude panning and simple 3-D sound processing for left-right spatialization. We used five different positions from left to right. The three middle positions are reproduced using amplitude panning. For the extreme left and right position we used a

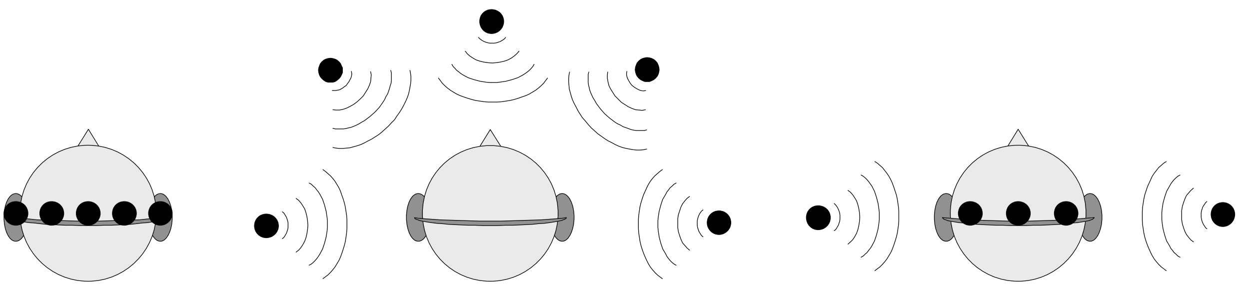

Fig. 1. Typical sound perception experienced over headphones for five sound positions obtained with stereo panning (left picture), HRTF processing (middle picture) and mixed presentation (right picture).

stereo-widening network developed by Kirkeby [7] to produce a pleasant out-of-head localization.

In the stereophonic sonification, we wanted to test the feasibility of spatial reproduction for the navigation, so we decided to use the position only to present the column number. Indeed, in the early design phase, this was found to work well. From the monophonic case, we knew the number of sound events to be a good candidate for presenting the row number. Both of these schemes provide absolute information on the position.

3.3. Other sonified parameters

In the current game implementation, the navigation does not allow a "wrap around" of position from the last column or row back to the first. Instead, a specific collision type sound is played, if the user tries to go over the limits of the grid. In the stereo design, this sound is spatialized similarly to the other navigation sounds according to the horizontal position.

The difference between cards already found, i.e., open cards and cards still to be found, is represented by a slight alteration of navigation sounds. A distortion was added to the unfound-card sounds, which in return gives a "muted" characteristic to the already-found-card sounds. This feature facilitates search for cards that are still part of the problem and is comparable to the difference in presentation between found and unfound cards in visual game. Additionally, if the user presses the middle key on an already found card, the sound of the card is not played.

After finding a matching pair of sounds, an indication on the number of found vs. unfound pairs is presented to the user. This is achieved by playing a series of sounds with ascending pitch. The number of sounds in the series corresponds the number of pairs in the game and the *n* last sounds of the series sound different to indicate that *n* pairs are still to be found. Clean vs. distorted sounds were used here again. Additionally, for the stereo sonification, the number of found cards is presented on the left, and the number of unfound cards on the right.

3.4. Implementation for the Compaq iPAQ

The memory game was implemented on a Compaq iPAQ handheld computer. The implementation is based on standard audio file playback capabilities of the device. Naturally, this is not the most efficient way to implement either of the sonification designs, but it is very easy to test different designs using this implementation.

One problem with the selected device and the program implementation was that the device produces extra clicks in the beginning and the end of each sound sample played. These clicks are localized mainly in the middle of the stereo image. This may sometimes be misleading with the localization-based stereo version of the sonification.

The implementation was designed to record all user input with timestamps to a text file, which can be parsed for each test subject. If necessary, we could reproduce any of the games the user played.

The input method utilized for both the visual interaction version and for the two different sonified versions was the five-way digital button on the device, i.e., the stylus is not used except for starting the game, and selecting visualization or sonification options.

4. USER TEST DESIGN

The implementation on the iPAQ handheld grants us much freedom when designing the usability test. We tested both the stereo and the mono sonification on all subjects. The test procedure consisted of two phases: a supervised test phase, and an unsupervised extended test period. As described earlier, all the user actions were recorded during the whole test procedure.

4.1. The supervised test

In the supervised test, we had twelve engineers and trainees as subjects. Most of the subjects deal with audio in their everyday work, and all were familiar with the concepts of pitch, rhythm, and stereo panning. These subjects cannot therefore be considered a group of average users. However, they were naïve with respect to the eyes-free audio memory game, and the sonification designs.

The supervised test phase began with a training part. In this part we let the subjects play the game for five to ten minutes in the visual mode, i.e., they saw the grid of cards during this phase. The purpose of this phase was to familiarize the subjects with the device, the sounds to be matched during the game, as well as the five-way key used for input during the test. In the training part, the subjects played the game without navigation sonification. The grid size of the game was different from the actual test phase.

The test phase was divided into three sections. In the first two sections, the subjects played the game using the mono and stereo sonifications. Half of the subjects played the monophonic version first, and the other half started with the stereophonic version. The users were asked to play each of the sonified versions for ten to fifteen minutes. In practice, most of the subjects played two games using each of the sonifications. The game played in the tests contained four rows and five columns.

No description on the sonification schemes was provided beforehand or during the testing, in order to assess the instant usability of the sonified versions and subjects' ability to learn the sonifications.

The subjects answered several questions after each sonifications. A qualitative assessment was included to compare the workload of the task for the mono and stereo presentation. A five-point scale questionnaire was employed, consisting of three questions, namely *performance*, *effort* and *comfort* in mental representation of the grid system. Four specific questions were also added to test the subject's understanding of each sonification scheme. After the second sonification, we also asked for the subject's overall preference between the mono and stereo sonifications.

Finally, the subjects (except for two) played one game with visual feedback but without sonification. This was included to get an overall baseline for the sonified versions.

4.2. The unsupervised test

After the supervised phase, two subjects were chosen for a longer test. These subjects were fully familiarized with the game and could easily navigate in the grid, tell the found and unfound cards apart, and understood the number-of-pairs-found sonification. They could also operate the device to change the game settings, i.e. enable or disable the visual display or the sonification, and change the type of sonification between mono and stereo.

The subjects were allowed to take the device with them for ca. one day. During that time, they were asked to play the game as many times as they like, and try the different game mode options. The user interaction data from this phase was analyzed separately from the supervised phase data.

5. RESULTS

In the supervised test, we mainly looked at qualitative differences in the instant usability of the two sonifications. The questionnaire was our main tool in analyzing the results of this phase. However, as the game data was collected during the test, some indication about the learning effects present in the test could also be obtained. The second test phase provided us with various quantitative data, but this data does not generalize too well, because it was acquired using only two subjects.

5.1. Results of the supervised test

The supervised test was designed with two groups of subjects, with two different orders of presentation. Six subjects heard the mono sonification first, and the six other subjects heard the stereo sonification. In the following, we call these groups A and B, respectively.

Two thirds of all subjects (i.e., 8 people) preferred stereo presentation to mono. However, it should be noted that three of the four subjects who preferred the mono presentation belong to the group B (i.e. stereo sonification first). A possible habituation to the auditory presentation may explain this preference for the mono presentation.

An analysis of the answers given to the questionnaire about subjects' understanding of the two sonification schemes shows that 42 % of subjects were able to give the grid size exactly right (33 % for both mono and stereo presentation), others made mistakes by one. Two subjects found the exact grid size only for the second condition, which may be explained by the learning effect.

Nearly all subjects understood and could describe the sonification scheme used to represent the two dimensions of the grid (96 % for mono and 92 % for stereo). This is encouraging for the main purpose of this sonification, i.e. representing a given position in a small two-dimensional grid.

Only 25% of the subjects noticed the sonification scheme used to distinguish found and unfound pairs by modifying the navigation sounds, as described in section 3.3. However, all of those who did not notice this difference, commented that there should be a difference in the sounds, i.e., they were aware of the problem, but not of the intended solution. None of the subjects understood the sonification scheme used to present the number of pairs still to be found.

After analyzing the detailed questions regarding the sonifications, the qualitative grading of the workload were compared for the mono and the stereo presentation and between the two groups of subjects.

Question	Mono	Stereo	t-test
Performance	3.75	4.17	0.137
Effort	3.92	3.67	0.169
Comfort	3.33	3.58	0.215

Table 1. Qualitative assessment compared for the mono and stereo sonification schemes. Means for the 12 subjects are presented as well as the probabilities associated with a t-test.

Table 1 gives an indication that the stereo presentation is favored, i.e., it has higher performance and comfort and smaller effort. These results correlate well with the preference for this technique. However, these differences are not significant, due to the small number of subjects included in the test.

Interesting dissimilarities between groups can also be noted from the questionnaire results. There is a significant difference in performance between the groups A and B (*2-tailed t-test*, $p=0.04$ for mono, and $p=0.01$ for stereo, $df=23$). Table 2 gives the results individually for the two groups. Looking at the group A alone, we see that performance is almost significantly better for the stereo presentation. No significant difference was found for the group B.

	Question	Mono	Stereo	t-test
Group A	Performance	4.17	4.67	0.07
	Effort	4.17	4.00	0.61
	Comfort	3.67	3.83	0.77
Group B	Performance	3.33	3.67	0.53
	Effort	3.67	3.33	0.46
	Comfort	3.00	3.33	0.36

Table 2. Questionnaire results presentation individually for the two groups, means for each group of 6 subjects.

To check the learning that happened during the supervised test, we plotted the average times between two consecutive key-presses for each user and averaged for all users (see Fig 2). This measure reflects the effort needed and comfort in the navigation. The trend in Fig. 2 is very clear: people get faster as they learn the navigation. The learning curve is also relatively fast, considering that no description was provided before the test on the sonification schemes. It is hard to see however, if the learn-

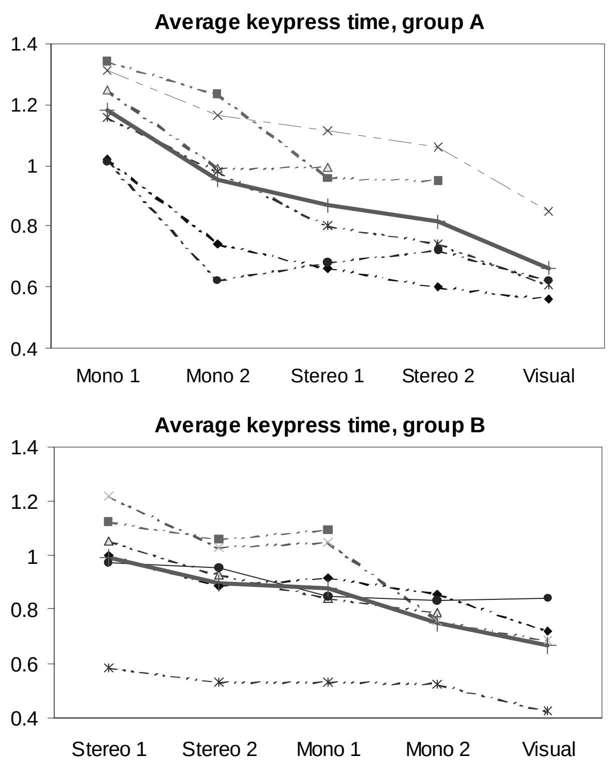

Fig. 2. The average time (in seconds) between two consecutive key-presses for each subject and in average.

ing has ended, when the subjects play their fourth game; the time between key-presses is still longer than in visual case, but this may be the case also after a longer training period, as can be seen from the results in the next section.

Few results would be significant by themselves from the first test phase with twelve subjects. However, all measures suggest a slightly better usability with stereo sonification.

5.2. Results of the unsupervised test

The two subjects participating in the second test phase played numerous games during the test period. Subject A played 14 visual, 13 mono, and 12 stereo games, and for subject B the number of games were 11, 13, and 12, respectively. Table 3 presents the average times (in seconds) needed to complete the game, the average number of cards opened during a game, the average total number of key-presses for one game, and the average times (in seconds) between two consecutive key-presses.

Subj.	Quantitative measure *		Condition		
			Visual	Mono	Stereo
A	Time (s)	Mean [1,2]	70.71	153.62	130.92
		St. error	2.43	7.06	6.32
	Cards open	Mean	39.29	48.15	42.00
		St. error	1.23	2.09	2.18
	Key presses	Mean [1,2]	123.00	175.23	147.75
		St. error	4.32	8.85	9.14
	Time/ event	Mean [1,2,3]	.577	0.880	0.903
		St. error	0.014	0.016	0.039
B	Time (s)	Mean [1,2]	96.73	157.08	189.08
		St. error	6.92	7.17	19.59
	Cards open	Mean [1]	48.91	59.54	57.67
		St. error	2.58	2.89	4.62
	Key presses	Mean [1]	152.73	225.23	231.33
		St. error	9.66	12.43	20.57
	Time/ event	Mean [1,2]	0.632	0.703	0.811
		St. error	0.013	0.013	0.027

* A paired-samples t-test was used to analyze the data. Significant differences (*p<0.01*) were found between:
[1] the visual and mono presentations,
[2] the visual and stereo presentations,
[3] the mono and stereo presentations.

Table 3. Quantitative measures of game performance for two subjects compared for the three conditions (i.e. visual and audio with mono or stereo sonification). Data presented includes averages over all games of the same type and standard errors.

From these results, it should be noted first that significant differences exist between the two subjects for the measures *time*, *cards opened* and *key presses* (*One-way ANOVA, p<0.01)*. However, a similar trend can be seen for the two subjects with significantly shorter game completion times for the visual navigation; the auralized versions of the game take almost twice as long to complete as the game based on visual navigation. Larger variations in completion times can also be seen for the two audio presentations.

This clear result was expected due to the serial nature of the auditory presentation, i.e. the user only hears information about his current position with the sonification approach proposed here, whereas information about all cards can be seen at one glance with the visual condition. However, looking now at the three other measures, we see that results for the number of open cards does not follow the same trend. Indeed, no significant differences were found for subject A, and differences were only significant between the mono and the visual presentation for subject B. This indicates that subjects tend to navigate more when searching for a specific position on the grid, which inevitably increases the overall time per game. However, in most of the cases, errors in finding the intended card are relatively low, which proves that the sonification of the grid works well.

Considering subjects separately now, we notice that subject A performed better with the stereo presentation than the monophonic one overall; however, only differences in average times between two consecutive key-presses (time/event) are significant. On the other hand, subject B seems to perform better with the mono sonification, though none of the differences are significant.

Finally, an interesting note can be made regarding the average times between key-presses. For subject A, the difference in this navigation speed between the visual and sonified games is quite large, but for subject B, the navigation speeds are more uniform. However, subject A seems to use the extra time between the key-presses to remember the location of the cards better.

6. CONCLUSIONS

In this paper, we have presented an audio-only memory game as a case study of auditory navigation in a small two-dimensional grid. The needs for sonification elements were described, and two different sonifications — one suitable for monophonic playback and another requiring stereo capabilities — were presented.

We organized the usability tests for the game in two parts: in the first phase we tried to assess the instant usability of the sonifications, and in the second phase we were interested in the long-term usability aspects.

The number of subjects who participated to the first test was too small to obtain statistical significance; however, results suggest slightly better performance for the stereo sonification than for the mono sonification. In addition, we found that testing the instant usability was very useful for the sonification design phase. Only one fourth of the subjects understood the difference between the sonification for the two different card statuses, which emphasizes the need to redesign that sonification scheme. The same applies to the sonification for the number of pairs still to be found that none of the subjects understood during the test.

In the second test phase, only two subjects tested the game, but the test period for both was quite long. The subjects were fully familiar with the sonifications before the test, so they can be considered expert users. As expected, the visual navigation is the most efficient way of playing the game. However, comparing results for the average number of open cards per game, we can conclude that eyes-free navigation in the small 2-D grid works relatively well for both the mono and stereo sonifications. There seem to be individual differences between the subjects regarding the performance in the sonified modes; for subject A stereo sonification is clearly more efficient, while the subject B seems to perform better using the monophonic version.

While the test application is a simple game, the sonification designs have significance also in other applications, e.g., two-dimensional icon based user interfaces, or navigation in small grids of tabular data.

7. ACKNOWLEDGMENTS

Special thanks to Juha Arrasvuori and his team for the original implementation of the audio memory game. This work has been supported by Nokia Foundation and Tekniikan edistämissäätiö.

8. REFERENCES

[1] Milton Bradley Co. “The Original Memory Game®” 1966, Div. of Hasbro, Inc.

[2] B. Uskali, “Sound Match” [Software]. Available from <http://www.abc.net.au/children/games/stories/s18309.htm>

[3] Hyro Pty Ltd. “The Sound Memory Game” [Online]. <http://www.hyro.com/games/pages/game_3/sound_game.swf>

[4] Kaiser-Kaplaner, “Music-Memo-Game” [Software] 1999. Available from <http://www.musica.at/shareware1/>

[5] R. Ramloll, S. Brewster, W. Yu, and B. Riedel, “Using Non-speech Sounds to Improve Access to 2D Tabular Numerical Information for Visually Impaired Users” In Proc. of BCS IHM-HCI 2001 (Lille, France), Springer, pp. 515–530.

[6] G. Lorho, J. Marila, and J. Hiipakka, “Feasibility of Multiple Non-Speech Sounds Presentation Using Headphones” in Proc. Int. Conf. on Auditory Display (ICAD) 2001 (Espoo, Finland), pp. 32–37.

[7] O. Kirkeby, “A Balanced Stereo Widening Network for Headphones” in Proc. AES 22nd Int. Conf. on Virtual, Synthetic and Entertainment Audio (Espoo, Finland).